

The Effect of External Representations On Numeric Tasks

Jiajie Zhang

Department of Psychology

The Ohio State University

1827 Neil Avenue

Columbus, OH 43210-1222

zhang.52@osu.edu

Tel: (614)-292-8667

Fax: (614)-292-5601

Hongbin Wang

Department of Psychology

The Ohio State University

1827 Neil Avenue

Columbus, OH 43210-1222

wang.190@osu.edu

ABSTRACT

This article explores the effect of external representations on numeric tasks. Through several trivial modifications on the previously reported two-digit number comparison task, we obtained completely different results. Rather than a holistic comparison, we found parallel comparison. We argue that this difference was a reflection of different representational forms: the comparison was based on internal representations in previous studies but on external representations in our present study. This representational effect was discussed under a framework of distributed number representations. We propose that in numerical tasks involving external representations, numbers should be considered as distributed representations and the behavior in these tasks should be considered as the interactive processing of internal and external information through the interplay of perceptual and cognitive processes. We suggest that theories of number representations and process models of numerical tasks should consider external representations as an essential component.

The Effect of External Representations On Numeric Tasks

Jiajie Zhang & Hongbin Wang

Department of Psychology
The Ohio State University

Many numerical tasks involve external representations. In these tasks, people need to process information distributed across internal and external representations in an interactive manner (see Zhang & Norman, 1995). For example, to multiply 735 by 278 using paper and pencil, we need to process not just the information in internal representations (e.g., the value of each individual symbol, the addition and multiplication tables, arithmetic procedures, etc.) but also the information in external representations (e.g., the visual and spatial properties of the symbols, the spatial relations of the partial products, etc.).

There are at least two different views on the roles of external representations in numerical tasks. One view is that external representations are merely input and stimuli to the internal mind. In this view, external representations first have to be re-represented as internal representations through some encoding processes. Only after this encoding stage does real numerical cognition take place: memorial and cognitive processes of numerical facts and procedures operate upon the internal number representations. At the end, the products of the numerical processing are externalized into the environment through some decoding processes. This is the view reflected in many studies on numerical cognition (see Dehaene, 1993, for a collection of reviews).

A different view is that external representations need not be re-represented as internal representations in order to be involved in numerical tasks: they can directly activate perceptual processes and directly provide perceptual information that, in conjunction with the memorial information and cognitive processes provided by internal

representations, determine the behavior in numerical tasks (e.g., Zhang & Norman, 1994, 1995; Zhang, in press). In this view, perceptual processes are not the peripheral encoding processes that re-represent external representations into internal representations. Rather, they directly operate upon external representations as the basic components of numerical tasks. Thus, in numerical tasks that involve external representations, the behavior is simply the integrative processing of the information perceived from external representations and that retrieved from internal representations through the interplay of perceptual and cognitive processes.

This article uses number comparison as an example to demonstrate the important roles of external representations in numerical tasks. We only make a few trivial modifications on the two-digit number comparison task that has been previously published (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981). The results, however, are quite different. They raise concerns over the existing theories of number representations. Their implications will be discussed under a framework of distributed number representations.

NUMBER COMPARISON

Arabic numerals have two dimensions: a base dimension represented by the shapes of the ten digits (0, 1, ..., 9) and a power dimension represented by the positions of the digits (for details, see Zhang & Norman, 1995). The base dimension is an internal representation because the numerical value of each digit (arbitrary shape) has to be memorized, whereas the power dimension is an external representation because the position of a digit can be perceptually inspected. The information needed to compare the magnitudes of two one-digit Arabic numerals is solely in internal representations because the comparison is only on the base dimension that is internal. In contrast, the information needed to compare two multidigit Arabic numerals is not only in internal representations but also in external representations: the values of individual digits on the base dimension

are stored in memory but the positions of individual digits on the power dimension are perceptually available in external representations. Thus, due to the different information sources, there might be different comparison mechanisms for one-digit and multidigit Arabic numerals.

For one-digit Arabic numerals, the time to make magnitude comparisons (larger or smaller) decreases linearly with the logarithm of the numerical distance between them (Moyer & Landauer, 1967; for reviews, see Banks, 1977; Moyer & Dumais, 1978). For example, it is faster to compare 1 and 9 than 8 and 9. This *distance effect* remains even when the comparison is based on English written number words (Foltz, Poltrock, & Potts, 1984) and Japanese kanji and kana numerals (Takahashi & Green, 1983). In addition, the distance effect resembles that found for physical stimuli such as line lengths and dot patterns (e.g., Buckley & Gillman, 1974). These results suggest that different types of numerals may have a common internal representation, which is similar to a physical continuum such as a format-independent *line-like* analog representation (e.g., Dehaene, Dupoux, and Mehler, 1990; Dehaene, 1992; Gallistel & Gelman, 1992; Restle, 1970).

For multidigit Arabic numerals, there are at least three different models of comparison (Hinrichs, Yurko, Hu, 1981). Let us use the reaction times of comparing a series of target numerals (e.g., 11 to 54 and 56 to 99) with a fixed standard numeral (e.g., 55) to describe the three models. The first model is sequential comparison: the two multidigit Arabic numerals are compared digit by digit sequentially. In this case, only the highest digits should affect the comparison unless they are not sufficient for making decisions. For example, the larger numeral of 29 and 55 can be decided by comparing their decade digits (2 and 5) alone. Thus, the RT for comparing 29 with 55 should be identical to the RT for comparing 21 with 55. Equation 1 is one likely sequential model for two-digit numerals, which is shown graphically in Figure 1A. D_s and D_t are the decade digits and U_s and U_t are the unit digits of the standard numeral and the target numeral, respectively; and a , b , and d are the parameters. When the decade digits of the

standard and the target are different ($D_s \neq D_t$), the RTs of the targets within a decade are identical to each other and the RTs for different decades decrease with the logarithm of the numeric distance between the decade digits of the standard and the target. When the decade digits of the standard and the target are identical ($D_s = D_t$), that is, the target is in the same decade as the standard, the RTs decrease with the logarithm of the numerical distance of the unit digits of the standard and the target. For both cases, the comparison is based on one-digit comparison.

The second model is parallel comparison: the lower digits may facilitate or interfere with the comparison of higher digits (i.e., a Stroop-like effect). For example, the unit digit 9 in 29 may increase the RT for comparing 29 with 55 and the unit digit 1 in 31 may decrease the RT for comparing 31 with 55. Equation 2 is one likely parallel model for two-digit numerals, which is shown graphically in Figure 1B. This model is identical to the sequential model shown in Equation 1 except that there is an extra term for the Stroop-like effect for targets whose decade digits are different from that of the standard. This Stroop-like term is a simple linear function of the numeric distance between the unit digits of the standard and the target. For targets smaller than the standard, this term has a facilitation effect for targets whose unit digits are smaller than that of the standard and an interference effect for targets whose unit digits are larger than that of the standard. For targets larger than the standard, the effects are reversed, as specified by the sign function in the equation.

The third model is holistic comparison: multidigit numerals are first encoded as integrated representations and then their whole numerical values are compared. In this case, only the absolute numerical values should matter. For example, the RTs for comparing target numerals 11, 32, .., and 54 with the standard 55 should be a smooth decreasing function of the numerical distances between the target numerals and the standard. Equation 3 is one likely holistic model for two-digit numerals, which is shown graphically in Figure 1C. It shows that the unit and decade digits of the target are first

integrated as a whole numerical value, which is then compared with the whole numerical value integrated from the decade and unit digits of the standard.

Empirical studies have revealed a discrepancy between two-digit comparison and higher multidigit (3 or more) comparison: two-digit comparison is holistic (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981) whereas higher multidigit comparison is sequential (Hinrichs, Berie, & Mosell, 1982; Poltrock & Schwartz, 1984). The holistic comparison for two-digit Arabic numerals is counter-intuitive because the decades of two two-digit Arabic numerals, if they are different, are sufficient to decide which numeral is larger or smaller. This finding has often been cited as evidence for a holistic analog internal representation of Arabic numerals.

In the experiments that follow, we show that whether the comparison of two-digit Arabic numerals is sequential, parallel, or holistic depends on how the two numerals are represented.

$$(1) \text{ Sequential Model } \begin{cases} RT = a - b \cdot \ln|D_s - D_t| & \text{if } D_s \neq D_t \\ RT = d - b \cdot \ln|U_s - U_t| & \text{if } D_s = D_t \end{cases}$$

$$(2) \text{ Parallel Model } \begin{cases} RT = a - b \cdot \ln|D_s - D_t| + c \cdot (U_t - U_s) \cdot \text{sign}(D_s - D_t) & \text{if } D_s \neq D_t \\ RT = d - b \cdot \ln|U_s - U_t| & \text{if } D_s = D_t \end{cases}$$

$$(3) \text{ Holistic Model } \quad RT = a - b \cdot \ln|(10 \cdot D_s + U_s) - (10 \cdot D_t + U_t)|$$

Insert Figure 1 about here

OVERVIEW OF EXPERIMENTS

In the experiments reported by Dehaene, Dupoux, and Mehler (1990) and Hinrichs, Yurko, and Hu (1981), a standard (e.g., 55) was always held in memory and only the target numerals (e.g., 11 to 99 except 55) were presented on an external display and judged whether smaller or larger than the standard. In this case, the comparison was between a preprocessed internal representation of the standard and an external representation of the target. In the first and second experiments of our current study, we made three changes. First, the standard and targets were presented simultaneously on a computer screen such that the comparison was between two external representations. Second, instead of one standard, we used two standards (55 and 65) as a within-subject factor to prevent subjects from preprocessing a specific standard and transforming it into an internal representation. Third, instead of reporting whether a target was smaller or larger than the standard, subjects only decided which of the two presented numerals was larger (Experiment 1) or smaller (Experiment 2). Such procedural changes are important for the testing of our hypothesis: when both numerals are presented as external representations, the comparison is not just based on internal but also on external information, the concurrent processing of which might generate a different pattern of behavior from that found with an internal and an external representation. To further test our hypothesis, we also conducted a control experiment (Experiment 3) in which the comparison was between an external representation and an internal representation.

There are over three thousand pairs of two-digit numerals. To make the experimental design manageable, the comparisons in our current study, as in previous studies, are between a set of target numerals and one or two fixed standards. For example, numerals 11 to 54 and 56 to 99, which are the target numerals, are compared with 55, which is the standard. In order to identify which of the three comparison models in Figure 1 can best describe two-digit comparison, we need four observations on the RTs of comparing target numerals with a standard. The first is the *unit effect*: the target

numerals within a decade (e.g., 21-29) have different RTs from each other. The second is the *decade effect*: the target numerals in a decade (e.g., 21-29) have a different average RT than that in a different decade (e.g., 31-39). The third is the *discontinuity effect*: there is a sharp change in RTs for target numerals across a decade boundary. For example, the RT difference between 29 and 30 can be much bigger than the RT difference between 28 and 29. The fourth is the *Stroop-like effect*: the unit digits may interfere or facilitate the comparison of the two-digit numerals. One difficulty with the observation of a Stroop-like effect is that the values of unit digits and the absolute distances from the target numerals to the standard are always confounded. For example, a faster RT to compare 21 with 55 than 29 with 55 can be either due to the facilitation of the 1 in 21 coupled with the interference of the 9 in 29, or the longer distance between 21 and 55 than between 29 and 55, or a combination of both. However, if we can observe a *reverse distance effect* across decade boundaries, then we can positively identify a Stroop-like effect. For example, the RTs for 26-29 might be slower than those for 31-34 even if 26-29 are farther away from 55 than 31-34 (see Figure 1B). Therefore, rather than testing a general Stroop-like effect, we will test the reverse distance effect in the experiments, which is a strong Stroop-like effect.

Based on the statistical significance of these four effects, we can evaluate the three models. Due to the nature of hypothesis testing, we can only confirm the existence of an effect by rejecting the null hypothesis when p is below the critical value. However, we cannot accept the null hypothesis and thereby confirm the nonexistence of an effect when p is above the critical value. Therefore, we can only use the existence of an effect to evaluate a model. First, if the comparison is sequential, then the decade effect and the discontinuity effect should be present but the unit effect and the reverse distance effect should not be present. Thus, if either the unit effect or the reverse distance effect is present, then sequential comparison can be rejected. Second, if the comparison is parallel, then the decade effect and the unit effect should be present and the discontinuity

effect may or may not be present.. The Stroop-like effect should also be present, but the stronger reverse distance effect may or may not be present. Because the Stroop-like effect is the defining feature of parallel comparison, its presence is sufficient for accepting the parallel model. Thus, if the reverse distance effect is present, then parallel comparison can be accepted. Third, if the comparison is holistic, then the unit effect and the decade effect should be present but the discontinuity effect and the reverse distance effect should not be present. Thus, if either the discontinuity effect or the reverse distance is present, then holistic comparison can be rejected.

In sum, by testing the four effects, we can only disprove (reject) the sequential model but cannot prove (accept) it. Similarly, we can only disprove the holistic model but cannot prove it. For the parallel model, in contrast, we can only prove it but cannot disprove it. Due to the limitations of the inferential statistics, we will use nonlinear regression to gather supplemental evidence for model selection by fitting the experimental data against each of the three model equations. The one that has the best fit of the data is the most likely model.

EXPERIMENT 1

In this experiment, the comparison is between two two-digit Arabic numerals that are both presented externally. We predict that unlike the comparison between an internal and an external two-digit Arabic numeral, the comparison under the current condition is no longer holistic. Target numerals 11 to 54 and 56 to 99 were compared with standard 55, and target numerals 31 to 64 and 66 to 99 were compared with standard 65. The symmetrical arrangement of target numerals below and above the standard makes it easier to observe whether the target numerals below and above the standard are compared in different ways. The task was to decide which of two numerals was larger.

Method

Subjects

The subjects were 32 undergraduate students in introductory psychology courses at The Ohio State University, who participated in the experiment to earn course credit. They were all native English speakers.

Design and procedure

The subjects were seated in about 40 cm from a Macintosh computer in a dark room. They were told that two Arabic numerals would appear on the screen simultaneously, one on the left and one on the right side of a fixation point. They were asked to press the left key ('z') or the right key ('m') as quickly and accurately as possible depending on whether the numeral on the left or the one on the right side is larger. The Macintosh computers (Quadra 700 and Centris 610), improved by a millisecond timer software¹, could measure reaction times with a resolution of ± 1 ms. Each pair of numerals were presented for 2 s, preceded by a fixation point (a '+' sign) of 500 ms and followed by a blank screen of 2 s. The two-digit Arabic numerals were in 24 point bold New York font (approximately 1.0 by 0.65 cm for each digit) and with an equal distance of 0.95 cm from the fixation point.

Arabic numerals 11 to 99 (target numerals) except 55 were compared with 55 and 31 to 99 (target numerals) except 65 were compared with 65. Thus, there were 156 target-standard pairs. Each target-standard pair was presented twice, with one trial having the standard on the left side and the other having the standard on the right side. Therefore, there were a total of 312 trials. These 312 trials were randomized as a whole block for each subject and then divided into four blocks with 78 trials for each block. No effort was made to prevent the same target-standard pair from appearing in the same block or on consecutive trials. Each subject was presented 10 randomly generated pairs

of two-digit numerals for practice, followed by the four blocks of experimental trials with one minute rests between blocks.

Results

For all analyses that follow, target-standard pairs with the standard on the left side were pooled with the same pairs with the standard on the right side. Trials with errors were excluded from the analysis of RTs. For standard 65, the average error rate was 2.7%, ranging from 1.5% in the 30s and 90s to 5.6% in the 60s. For standard 55, the average error rate was 2.0%, ranging from 0.64% in the 10s and 90s to 4.2% in the 50s. RTs that deviated from the mean for each target by more than three standard deviations were excluded from analyses. Separate analyses were conducted for standards 65 and 55.

Standard 65

The average RTs are shown in Figure 2. An analysis of the effect of the numerical distances between targets and 65 and the ranges of the targets (smaller or larger than 65) showed a significant distance effect ($F(33, 693) = 16.99, p < 0.001$), a significant range effect ($F(1, 21) = 8.40, p < 0.009$; 710.5 ms and 687.7 ms for targets below and above 65), and a significant interaction ($F(33, 693) = 4.86, p < 0.001$).

Decade and unit effects. A three-way ANOVA on mean RTs for ranges (below or above 65), decades (30s-50s and 70s-90s, excluding 60s), and units (1-9) showed an insignificant range effect ($F(1, 23) = 0.23, p = 0.64$), a significant decade effect ($F(2, 46) = 7.09, p < 0.005$), and an insignificant unit effect ($F(8, 184) = 0.154, p = 0.15$). The two-way interactions were significant between ranges and units ($F(8, 184) = 2.22, p < 0.05$), but not between ranges and decades ($F(2, 46) = 1.90, p = 0.16$) and between units and decades ($F(16, 368) = 1.22, p = 0.25$). The three-way interaction was not significant ($F(16, 368) = 1.26, p = 0.22$). Separate analyses were carried out for targets below and above 65. For targets 31-59, a two-way ANOVA on mean RTs for decades (30s, 40s, and 50s) and units (1-9) showed a significant decade effect ($F(2, 50) = 60.75, p < 0.001$), a

significant unit effect ($F(8, 200) = 16.15, p < 0.001$), and a significant interaction ($F(16, 400) = 2.71, p < 0.001$). There was a significant decade effect between 40s and 50s: the RTs for 40s were faster than those for 50s ($F(1, 25) = 77.93, p < 0.001$). However, there was no significant decade effect between 30s and 40s: the RTs did not differ between 30s and 40s ($F(1, 27) = 0.94, p = 0.34$). For each of the decades of 30s, 40s, and 50s, the unit effect was significant (smallest $F(8, 216) = 3.91, p < 0.001$). For targets 71-99, the decade effect was significant ($F(2, 46) = 20.18, p < 0.001$) but the unit effect was not significant ($F(8, 184) = 0.31, p = 0.96$). The interaction was not significant ($F(16, 368) = 1.13, p = 0.32$).

The unit effect was further analyzed by linear regression. The RT for each target in a decade was subtracted by the mean RT of the corresponding decade, then averaged across 30s, 40s, & 50s and across 70s, 80s, & 90s (see Figure 4A). The unit effect was asymmetrical. For targets smaller than 65, the units had a strong effect with a slope of 13.7, which was significantly different from zero ($R^2 = 0.71, p = 0.005$). However, for targets larger than 65, the units had no significant effect: the slope (-0.66) was not significantly different from zero ($R^2 = 0.024, p = 0.69$).

Discontinuity Effect. If there is a discontinuity at the boundary of two decades, there should be a sharp change in RT. Adopting a calculation similar to that used by Dehaene, Dupoux, and Mehler (1990), the change in RT across a decade boundary (e.g., $RT_{69} - RT_{70}$) was compared with the averaged change in RT between adjacent numbers within each of the two adjacent decades (e.g., $[(RT_{68} - RT_{69}) + (RT_{70} - RT_{71})]/2$). An analysis of variance showed significant discontinuity effects between all decades (smallest $F(1, 30) = 4.76, p < 0.05$) except of between 40s and 50s ($F(1, 30) = 0.32, p = 0.57$).

Reverse Distance Effect. For targets below 65, there was a reverse distance effect across the boundary between 30s and 40s: RTs for 36-39 were significantly slower

than those for 41-44 ($F(1, 9) = 20.80, p < 0.001$). For targets above 65, no reverse distance effect was found.

Model Fitting. The average RTs shown in Figure 2 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above the standard 65. R^2 , an index of how much of the variance in the data can be accounted for by the model, was obtained for each model, as shown in Table 1 in the parentheses. Because the three models have different numbers of parameters (2, 3, and 4 for holistic, sequential, and parallel models, respectively), we cannot directly use R^2 to decide which model has the best fit. To take the number of parameters into account, we used the Akaike Information Criterion² (AIC), which is an index of fit that penalizes more heavily models with more parameters as opposed to those with fewer parameters (Akaike, 1973, 1983; see also Myung & Pitt, in press). The smaller the AIC of a model, the better the fit. The AIC values (Table 1) indicate that the parallel model is the best fit for targets below the standard and the sequential model is the best fit for targets above the standard.

Insert Figure 2 about here

Insert Figure 3 about here

Insert Figure 4 about here

Insert Table 1 about here

Standard 55

The average RTs are shown in Figure 3. There was a marginally significant range effect ($F(1, 16) = 3.84, p = 0.07$; 652.0 ms and 665.5 ms for targets below and above 55), a significant distance effect ($F(43, 688) = 15.23, p < 0.001$), and a significant interaction between range and distance ($F(43, 688) = 2.25, p < 0.001$).

Decade and Unit Effects. A three-way ANOVA on mean RTs for ranges (below or above 55), decades (10s-40s and 60s-90s, excluding 50s), and units (1-9) showed an insignificant range effect ($F(1, 17) = 0.66, p = 0.43$), a significant decade effect ($F(3, 51) = 65.89, p < 0.001$), and a significant unit effect ($F(8, 136) = 7.55, p < 0.001$). The two-way interactions were all significant (largest $p < 0.05$). The three-way interaction was not significant ($F(24, 408) = 0.69, p = 0.87$). Separate analyses were carried out for targets below and above 55. For targets 11-49, an ANOVA on mean RT for decades (10s, 20s, 30s, 40s) and units (1-9) showed a significant decade effect ($F(3, 69) = 39.02, p < 0.001$) and a significant unit effect ($F(8, 184) = 8.19, p < 0.001$). The interaction was not significant ($F(24, 552) = 1.34, p = 0.13$). For targets 61-99, there was a significant decade effect ($F(3, 60) = 74.77, p < 0.001$), a significant unit effect ($F(8, 160) = 2.42, p < 0.05$), and a significant interaction ($F(24, 480) = 1.57, p < 0.05$). The decade effect was significant between 60s and 70s and between 70s and 80s (smallest $F(1, 22) = 27.66, p < 0.001$), but not between 80s and 90s ($F(1, 24) = 0.50, p = 0.49$). The unit effect was significant for 60s ($F(8, 232) = 2.32, p = 0.02$) but not for 70s, 80s, and 90s (largest $F(8, 208) = 1.41, p = 0.19$).

Using the same method as for standard 65, the unit effect for standard 55 was also analyzed by linear regression (see Figure 4B). The unit effect was asymmetrical. For targets smaller than 55, the units had a strong effect with a slope of 8.75, which was significantly different from zero ($R^2 = 0.81, p < 0.001$). However, for targets larger than 55, the units had no significant effect: the slope (-2.25) was not significantly different from zero ($R^2 = 0.33, p = 0.10$). This is different from the significant unit effect from

the above ANOVA analysis. It might be because the linear regression was carried out on the averaged data.

Discontinuity Effect. Using the same discontinuity test for standard 65, it was shown that there was a marginally significant discontinuity effect between 50s and 60s ($F(1, 30) = 3.80, p = 0.06$) and a significant discontinuity effect between 60s and 70s ($F(1, 30) = 6.75, p < 0.01$). There was no significant discontinuity effect at other decade boundaries (largest $F(1, 30) = 2.91, p = 0.10$).

Reverse Distance Effect. For targets both below and above 55, no reverse distance effect was found.

Model Fitting. Similar to the analysis for standard 65, the average RTs shown in Figure 3 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above 55. The AIC (Table 1) indicates that the parallel model is the best fit for targets both below and above 55, although for targets above 55 the AIC value for the holistic model is only slightly larger than that for the parallel model.

Summary and Discussion

For targets below 65, the unit effect, the decade effect, the discontinuity effect, and the reverse distance effect were all significant. The significance of the reverse distance effect indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential models should be rejected. And the significance of both the reverse distance effect and the discontinuity effect indicates that the holistic model should be rejected. The AIC also indicates that the parallel model had the best fit of the data. For targets above 65, the decade effect and the discontinuity effect were significant but the unit effect and the reverse distance effect were not significant. The significance of the discontinuity effect indicates that the holistic model should be rejected. The parallel and the sequential models could neither

be rejected nor accepted. However, the AIC indicates that the sequential model has the best fit of data. Thus, sequential comparison was more likely than parallel comparison for targets above 65.

For targets both below and above 55, the unit effect, the decade effect, and the discontinuity effect were significant but the reverse distance effect was not significant. The significance of the unit effect indicates that the sequential model should be rejected. The significance of the discontinuity effect indicates that the holistic model should be rejected. The parallel model can neither be rejected nor accepted. However, the AIC indicates that the parallel model has the best fit of the data for targets both below and above 55. Thus, parallel comparison was the most likely model for targets both below and above 55.

In sum, this experiment showed the following results. First, for both standards, the holistic model was rejected for targets both below and above the standards. Second, for both standards, the sequential model was rejected and the parallel model was accepted for targets below the standards. Third, for both standards, parallel and sequential comparisons were neither rejected nor accepted for targets above the standards, although sequential comparison was more likely than parallel comparison for standard 65 and parallel comparison was more likely than sequential comparison for standard 55.

These results are completely at odd with those found by Dehaene, Dupoux, and Mehler (1990) and Hinrichs, Yurko, and Hu (1981), which showed a holistic comparison for two-digit numerals. To verify our results, we conducted another experiment to replicate the findings of Experiment 1. The replication experiment was identical to Experiment 1 except that 65 and 45 were used as the two standards. To save space, we only give the summary of the results. The replication experiment generated nearly identical results. First, the holistic model was rejected for targets below and above both standards. Second, the sequential model was rejected but the parallel model was accepted for targets below both standards. Third, for targets above both standards, the

sequential and the parallel models were neither rejected nor accepted, although the AIC indicates that the sequential model has a slightly better fitting than the parallel model for both standards.

EXPERIMENT 2

In Experiment 1, the task was to decide which of the two externally presented numerals was larger. The result is that the comparison was somehow asymmetrical with regard to parallel and sequential comparisons. For targets below the standards, parallel comparison was the only correct model. For targets above the standards, however, sequential and parallel comparisons were both likely. This asymmetry might be due to the asymmetry of the targets, that is, the comparison is always parallel below the standard but it could be either sequential or parallel above the standard. Another possibility is that the asymmetry might be due to the asymmetry of the task. When the task is to decide which numeral is larger, the comparison is always parallel below the standard but it could be either sequential or parallel above the standard. When the task is to decide which numeral is smaller, in contrast, the comparison could be either sequential or parallel below the standard but it is always parallel above the standard. A third possibility is that the asymmetry was merely an illusory phenomenon: the comparison is always parallel regardless of the whether the targets are below or above the standard and regardless of whether the task is to decide which numeral is larger or to decide which numeral is smaller. The probable sequential comparison above the standard in Experiment 1 might be merely a degraded parallel comparison. This is because the reverse distance effect we tested is a strong version of the Stroop-like effect. It could be that the Stroop-like effect was significant but it was not strong enough to be observed as a reverse distance effect.

Experiment 2 examines these three possibilities by asking subjects to decide which numeral is smaller as opposed to asking subjects to decide which numeral is larger in Experiment 1. If the comparison is always parallel below the standard but it could be

either sequential or parallel above the standard, then the hypothesis of target asymmetry is supported. If the comparison could be either sequential or parallel below the standard but it is always parallel above the standard, then the hypothesis of task asymmetry is supported. If the comparison is always parallel both below and above the standard, then the hypothesis of illusory asymmetry is supported.

Method

The design and procedure were exactly the same as in Experiment 1 except that instead of reporting which numeral was larger, the subjects reported which numeral was smaller. There were 32 subjects from the same subject pool as in Experiment 1.

Results

For all analyses that follow, trials with a standard on the left side were pooled with corresponding trials with the same standard on the right side. Trials with errors were excluded from the analysis of reaction times. For standard 65, the average error rate was 3.5%, ranging from 2.1% in the 30s and 90s to 5.7% in the 60s. For standard 55, the average error rate was 3.1%, ranging from 1.3% in the 10s and 90s to 5.2% in the 50s. RTs deviated from the mean for each target by more than three standard deviations were excluded from analyses. Separate analyses were conducted for 65 and 55.

Standard 65

The average RTs are shown in Figure 5. An analysis of the effect of the numerical distances between targets and 65 and the ranges of the targets (smaller or larger than 65) showed a significant distance effect ($F(33, 693) = 12.28, p < 0.001$), a significant range effect ($F(1, 21) = 5.82, p = 0.02$; 724.4 ms and 747.5 ms for targets below and above 65), and a significant interaction ($F(33, 693) = 6.43, p < 0.001$).

Decade and Unit Effects. A three-way ANOVA on mean RTs for ranges (below or above 65), decades (30s-50s and 70s-90s, excluding 60s), and units (1-9)

showed an insignificant range effect ($F(1, 22) = 5.71, p < 0.05$), a significant decade effect ($F(2, 44) = 36.25, p < 0.001$), and a significant unit effect ($F(8, 176) = 8.55, p < 0.001$). The two-way interactions were all significant (largest $p < 0.05$). The three-way interaction was also significant ($F(16, 352) = 2.06, p < 0.01$). Separate analyses were carried out for targets below and above 65. For targets 31-59, a two-way ANOVA on mean RTs for decades (30s, 40s, and 50s) and units (1-9) showed a significant decade effect ($F(2, 44) = 118, p < 0.001$), a significant unit effect ($F(8, 176) = 11.58, p < 0.001$), and a significant interaction ($F(16, 352) = 2.75, p < 0.001$). The decade effect was significant between 30s and 40s ($F(1, 25) = 7.95, p = 0.009$) and between 40s and 50s ($F(1, 22) = 114, p < 0.001$). For every decade below 65, the unit effect was significant (smallest $F(8, 224) = 3.29, p < 0.001$). For targets 71-99, the decade effect was not significant ($F(2, 50) = 0.82, p = 0.45$) but the unit effect was significant ($F(8, 200) = 3.51, p < 0.001$). The interaction was not significant ($F(16, 400) = 1.12, p = 0.25$).

As in Experiment 1, the unit effect for standard 65 was further analyzed by linear regression on the mean RTs averaged across subjects (see Figure 7A). The unit effect was asymmetrical. For targets smaller than 65, the units had a strong effect with a slope of 11.90, which was significantly different from zero ($R^2 = 0.58, p = 0.02$). However, for targets larger than 65, the units had no significant effect: the slope (-3.04) was not significantly different from zero ($R^2 = 0.10, p = 0.42$). This is different from the significant unit effect from the above ANOVA analysis. It might be because the linear regression was carried out on the averaged data.

Discontinuity Effect. Using the same discontinuity test as in Experiment 1, it was shown that there was significant discontinuity effect between 50s and 60s ($F(1, 25) = 8.85, p = 0.006$) and 60s and 70s ($F(1, 29) = 9.37, p = 0.005$) but not at other decade boundaries (largest $F(1, 29) = 1.75, p = 0.20$).

Reverse Distance Effect. For targets below 65, there were reverse distance effects across the boundaries between 30s and 40s and between 50s and 60s: RTs for 36-

39 were marginally slower than those for 41-44 ($F(1, 28) = 3.31, p = 0.08$) and RTs for 56-59 were marginally slower than those for 61-64 ($F(1, 26) = 3.98, p = 0.06$). For targets larger than 65, there was a reverse distance effect across the boundary between 70s and 80s: RTs for 76-79 were significantly faster than those for 81-84 ($F(1, 28) = 6.82, p = 0.01$).

Model Fitting. The same nonlinear regression analyses in Experiment 1 were carried out for the present experiment. The average RTs shown in Figure 5 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above the standard 65. For targets below 65, the parallel model is the best fit with smallest AIC (Table 1). For targets above 65, both the sequential and parallel models have better fit than the holistic model but the sequential model has a slightly better fit than the parallel model.

Insert Figure 5 about here

Insert Figure 6 about here

Insert Figure 7 about here

Standard 55

The average RTs are shown in Figure 6. There was a significant range effect ($F(1, 24) = 46.50, p < 0.001$; 649.6 ms and 689.1 ms for targets below and above 55), a significant distance effect ($F(43, 1032) = 12.48, p < 0.001$), and a significant interaction ($F(43, 1032) = 2.03, p < 0.001$). Due to the significant interaction, separate analyses were conducted for targets below and above 55.

Decade and Unit Effects. A three-way ANOVA on mean RTs for ranges (below or above 55), decades (10s-40s and 60s-90s, excluding 50s), and units (1-9) showed a significant range effect ($F(1, 20) = 39.94, p < 0.001$), a significant decade effect ($F(3, 60) = 30.83, p < 0.001$), and a significant unit effect ($F(8, 160) = 8.90, p < 0.001$). The two-way interactions were all significant (largest $p < 0.05$). The three-way interaction was not significant ($F(24, 480) = 0.69, p = 0.83$). Separate analyses were carried out for targets below and above 55. For targets 11-49, a two-way ANOVA for decades (10s, 20s, 30s, and 40s) and units (1-9) showed a significant decade effect ($F(3, 87) = 29.40, p < 0.001$) and a significant unit effect ($F(8, 232) = 5.13, p < 0.001$). The interaction was not significant ($F(24, 696) = 1.09, p = 0.34$). For targets 61-99, there were a significant decade effect ($F(3, 84) = 19.60, p < 0.001$), a significant unit effect ($F(8, 224) = 5.69, p < 0.001$), and a significant interaction ($F(24, 672) = 1.56, p = 0.04$). The decade effect was significant between 60s and 70s ($F(1, 27) = 23.42, p < 0.001$) and between 80s and 90s ($F(1, 25) = 4.60, p = 0.04$) but not between 70s and 80s ($F(1, 28) = 0.33, p = 0.57$). The unit effect was significant for 60s, 70s, and 80s (smallest $F(8, 240) = 2.08, p = 0.04$), but not for 90s ($F(8, 216) = 1.22, p = 0.29$).

As in Experiment 1, the unit effect for standard 55 was further analyzed by linear regression on the mean RTs averaged across subjects (see Figure 7B). For targets smaller than 55, the units had a strong effect with a slope of 6.40, which was significantly different from zero ($R^2 = 0.70, p = 0.005$). For targets larger than 55, the slope (-4.73) was marginally different from zero ($R^2 = 0.35, p = 0.09$).

Discontinuity Effect. Using the same discontinuity test as in Experiment 1, it was shown that there was no significant discontinuity effect at any decade boundaries (largest $F(1, 30) = 2.54, p = 0.12$).

Reverse Distance Effect. For targets below 55, there was a reverse distance effect between 20s and 30s: RTs for 26-29 were significantly slower than those for 31-34 ($F(1, 29) = 3.48, p = 0.02$). For targets larger than 55, there was also a reverse distance

effect between 70s and 80s: the RTs for 76-79 were significantly slower than those for 81-84 ($F(1, 26) = 6.77, p = 0.07$).

Model Fitting. The average RTs shown in Figure 6 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above the standard 55. For targets both below and above 55, the parallel and holistic models have a better fit than the sequential model but the parallel model has a better fit than the holistic model (Table 1).

Summary and Discussion

For targets below 65, the unit effect, the decade effect, the discontinuity effect, and the reverse distance effect were all significant. The significance of the reverse distance effect indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential models should be rejected. And the significance of both the reverse distance effect and the discontinuity effect indicates that the holistic model should be rejected. The AIC also indicates that the parallel model has the best fit of the data. For targets above 65, the unit effect, the discontinuity effect, and the reverse distance effect were significant but the decade effect was not significant. The significance of the reverse distance indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential model should be rejected. And the significance of both the reverse distance effect and the discontinuity effect indicates that the holistic model should be rejected. The AIC indicates that both the sequential and parallel models have a better fit than the holistic model. However, the sequential model has a slightly better fit than the parallel model (304 vs. 306 for the sequential and parallel models, respectively). Because the reverse distance effect is a sufficient feature for accepting the parallel model, its presence indicates that the comparison is parallel, not sequential.

For targets both below and above 55, the unit effect, the decade effect, and the reverse distance effect were significant but the discontinuity effect was not significant. The significance of the reverse distance effect indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential models should be rejected. And the significance of the reverse distance effect indicates that the holistic model should be rejected. The AIC also indicates that the parallel model has the best of the data for targets both below and above 55.

In sum, for both standards, the comparison was neither holistic nor sequential but parallel, both for targets below the standards and for targets above the standards. This result is neither consistent with the target asymmetry hypothesis that predicted parallel comparison below the standards and either sequential or parallel comparison above the standards, nor consistent with the task asymmetry hypothesis that predicted either sequential or parallel comparison below the standards and parallel comparison above the standards. However, it is consistent with the illusory asymmetry hypothesis that predicted parallel comparisons for targets both below and above the standards. This suggests that the apparent sequential comparison for targets above 65 was actually a degraded parallel comparison because the Stroop-like effect was not strong enough to be observed as a reverse distance effect.

EXPERIMENT 3

Combining the results of Experiments 1 and 2, we can reach the following conclusion: when two-digit number comparison is between two external representations, the comparison is parallel, not holistic. This is completely different from the result of the previous studies (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981) in which the comparison was found to be holistic when the comparison was between an internal and an external representation. There are two major differences between our present study and the previous studies. The first is the difference in representation forms:

two external representations in our present study but one internal and one external representation in the previous studies. The second is the difference in task demands. The previous studies used a classification task: the task was to decide whether the target was smaller or larger than the standard. In contrast, our current study used a selection task: the task was to decide which of the two numerals was larger or smaller. This difference in task demands was unlikely the cause of the different results because it has been shown that these two types of tasks are essentially equivalent (Dehaene, 1989). However, to directly test our hypothesis that the different comparison processes were due to different representations, we conducted Experiment 3 as a control experiment. In this experiment, the task was identical to the task in Experiments 1 and 2 of the present study, that is, it was a selection task. However, the comparison was between an internal and an external representation. If our hypothesis is correct, then we should find holistic comparison in this control experiment.

Method

The design and procedure were the same as in Experiment 1, except of the following changes. In Experiment 1, the comparison was between an external standard and an external target numeral, which were simultaneously presented on the left and right sides of a fixation point. In the current experiment, however, the comparison was between an internal standard and an external target numeral. In this new design, a standard (55 or 65) was presented first for one second either on the left side or the right side of the fixation point, followed by a three second blank interval for memory retention, and then followed by a target numeral on the other side of the fixation point. The three second blank interval between the standard and the target was to give subjects enough time to encode the standard as an internal representation. Subjects were told to make their responses as soon as the target appeared. There were 31 subjects from the same subject pool as in Experiment 1.

Results

For all analyses that follow, trials with a standard on the left side were pooled with corresponding trials with the same standard on the right side. Trials with errors were excluded from the analysis of reaction times. For standard 65, the average error rate was 4.1%, ranging from 2.9% in the 30s and 90s to 8.0% in the 60s. For standard 55, the average error rate was 3.0%, ranging from 2.2% in the 10s and 90s to 7.5% in the 50s. RTs deviated from the mean for each target by more than three standard deviations were excluded from analyses. Separate analyses were conducted for 65 and 55.

Standard 65

The average RTs are shown in Figure 8. An analysis of the effect of the numerical distances between targets and 65 and the ranges of the targets (smaller or larger than 65) showed a significant distance effect ($F(33, 627) = 9.38, p < 0.001$), a significant range effect ($F(1, 19) = 17.23, p < 0.001$; 584.6 ms and 550.4 ms for targets below and above 65), and a significant interaction ($F(33, 627) = 1.66, p = 0.01$). .

Decade and Unit Effects. A three-way ANOVA on mean RTs for ranges (below or above 65), decades (30s-50s and 70s-90s, excluding 60s), and units (1-9) showed a significant range effect ($F(1, 21) = 16.99, p < 0.001$), a significant decade effect ($F(2, 42) = 30.64, p < 0.001$), and an insignificant unit effect ($F(8, 168) = 0.95, p = 0.48$). The two-way interactions were significant between ranges and decades ($F(2, 42) = 8.87, p < 0.001$) and between ranges and units ($F(8, 168) = 2.01, p < 0.05$) but not between decades and units ($F(16, 336) = 1.31, p = 0.19$). The three-way interaction was marginally significant ($F(16, 336) = 1.58, p = 0.07$). . Separate analyses were carried out for targets below and above 65. For targets 31-59, a two-way ANOVA on mean RTs for decades (30s, 40s, and 50s) and units (1-9) showed a significant decade effect ($F(2, 46) = 27.40, p < 0.001$), a significant unit effect ($F(8, 184) = 2.11, p < 0.05$), and a marginally significant interaction ($F(16, 368) = 1.51, p = 0.09$). The decade effect was significant

between 40s and 50s ($F(1, 24) = 37.40, p < 0.005$) but not between 30s and 40s ($F(1, 12) = 0.18, p = 0.68$). The unit effect was significant for 50s ($F(8, 200) = 2.06, p < 0.05$) but not for 40s and 50s (largest $F(8, 208) = 1.16, p = 0.32$). For targets 71-99, the decade effect was significant ($F(2, 48) = 6.17, p < 0.005$) but the unit effect was not significant ($F(8, 192) = 1.49, p = 0.16$). The interaction was not significant ($F(16, 384) = 0.93, p = 0.54$).

As in Experiment 1, the unit effect was further analyzed by linear regression on the mean RTs (see Figure 10A). The unit effect was marginally significant for targets below 65 (slope = 1.94, $R^2 = 0.35, p = 0.09$) but not significant for those above 65 (slope = 1.34, $R^2 = 0.20, p = 0.22$).

Discontinuity Effect. Using the same discontinuity test as in Experiment 1, it was shown that there was no significant discontinuity effect at any decade boundaries (largest $F(1, 28) = 2.65, p = 0.12$).

Reverse Distance Effect. For target both below and above 65, no reverse distance effect was found.

Model Fitting. The average RTs shown in Figure 8 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above the standard 65 (Table 1). For targets below 65, the holistic and parallel have a better fit than the sequential model, and the holistic model has a slightly better fit than the parallel model. For targets above 65, the parallel model has the best fit.

Insert Figure 8 about here

Insert Figure 9 about here

Insert Figure 10 about here

Standard 55

The average RTs are shown in Figure 9. An analysis of the effect of the numerical distances between targets and 55 and the ranges of the targets (smaller or larger than 55) showed a significant distance effect ($F(43, 774) = 9.51, p < 0.001$) and a significant range effect ($F(1, 18) = 17.37, p < 0.001$; 560.0 ms and 526.6 ms for targets below and above 55). The interaction was not significant ($F(43, 774) = 1.09, p = 0.32$).

Decade and Unit Effects. A three-way ANOVA on mean RTs for ranges (below or above 55), decades (10s-40s and 60s-90s, excluding 50s), and units (1-9) showed a significant range effect ($F(1, 19) = 18.39, p < 0.001$), a significant decade effect ($F(3, 57) = 3.92, p < 0.01$), and a marginally significant unit effect ($F(8, 152) = 1.04, p = 0.07$). None of the two-way interactions were significant (smallest $p = 0.32$). The three-way interaction was not significant ($F(24, 456) = 0.99, p = 0.48$). Separate analyses were carried out for targets below and above 55. For targets below 55, a two-way ANOVA on mean RTs for decades (10s, 20s, 30s, and 40s) and units (1-9) showed a significant decade effect ($F(3, 69) = 6.03, p = 0.001$), a marginally significant unit effect ($F(8, 184) = 1.68, p = 0.10$), and an insignificant interaction ($F(24, 552) = 1.09, p = 0.35$). For targets above 55, there was a significant decade effect ($F(3, 69) = 2.98, p < 0.05$) and a significant unit effect ($F(8, 184) = 2.23, p < 0.05$). The interaction was not significant ($F(24, 552) = 1.06, p = 0.39$).

The unit effect was further analyzed by linear regression on the mean RTs (see Figure 10A). The unit effect was significant for targets smaller than 55 (slope = 2.34, $R^2 = 0.44, p = 0.05$) but not for those larger than 55 (slope = 0.29, $R^2 = 0.012, p = 0.78$).

Discontinuity Effect. Using the same discontinuity test as in Experiment 1, it was shown that there was no significant discontinuity effect at any of the decade boundaries (largest $F(1, 30) = 1.80, p = 0.19$).

Stroop-like Effect. For targets both below and above 55, no reverse distance effect was found.

Model Fitting. The average RTs shown in Figure 9 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above the standard 55 (Table 1). For targets below 55, the parallel model has the best fit. For targets above 55, the holistic and parallel models have a better fit than the sequential model, and the holistic model has a slightly better fit than the parallel model.

Summary and Discussion

For targets below 65, the unit effect and the decade were significant but the reverse distance effect and the discontinuity effect were not significant. The significance of the unit effect indicates that the sequential model should be rejected. However, the parallel and the holistic models can neither be rejected nor accepted. The AIC indicates that the holistic model has a slightly better fit than the parallel model. Thus, holistic comparison was the mostly likely model for targets below 65. For targets above 65, only the decade effect was significant. This indicates that none of the three models can be rejected or accepted. The AIC indicates that the parallel model has the best fit of the model. Thus, parallel comparison was the most likely model for targets above 65.

For targets both below and above 55, the unit effect and decade effect were significant but the reverse distance effect and the discontinuity effect were not significant. The significance of the unit effect indicates that the sequential model should be rejected. However, the parallel and the holistic models can neither be rejected nor accepted. The AIC indicates that the parallel model has the best fit of the data for targets below 55 and the holistic model has the best fit of the data for targets above 55. Thus, parallel comparison was the most likely model for targets below 55 and holistic comparison was the most likely model for target above 55.

The comparison in this experiment was between an internal and an external representation. We predicted that the comparison would be holistic. Although the results

did not converge strongly to support a holistic comparison model, they showed signs of holistic comparison. For both standards, sequential comparison was rejected but parallel and holistic comparisons could neither be rejected nor accepted. The model fitting results indicate that the holistic model was the best fit for half of the data and the parallel model was the best for the other half of the data. Therefore, at least for half of the data, holistic comparison was the most likely model. The parallel comparison for the other half of the data might be due to the incomplete internalization of the standards, which randomly alternated between 65 and 55 and were processed for internalization for a short period of three seconds. In the previous studies that showed holistic comparison, there was only one standard that was always in memory throughout the experiment.

One conclusion that we can tentatively, though not definitely, reach is that the difference between the holistic comparison found in the previous studies and the parallel comparison found in Experiments 1 and 2 of our current study was not due to the difference between the classification task used in the previous studies and the selection task used in our present study. This is because the current experiment showed that holistic comparison was also found in the selection task.

GENERAL DISCUSSION

The Effect of Representations

The experiments of the present study were trivial modifications of the previously published experiments. However, the results were quite different. In the previous studies by Dehaene, Dupoux, and Mehler (1990) and Hinrichs, Yurko, and Hu (1981), the comparison was holistic. In Experiments 1 and 2 of our present study, however, the comparison was no longer holistic: it was parallel. There are two differences between the previous and current studies: one in representations and one in task demands. Experiment 3 of the present study showed that the different comparison processes were not likely due to the differences in task demands.

Therefore, we argue that the difference in results between the previous and the present studies is a reflection of the difference in representations. In the previous studies, the standard was always in memory as an internal representation but the target was on the screen as an external representation. The standard in memory was no longer an Arabic numeral: it was converted to a phonological number word for memory retention. Because numerals can be compared only when they are in the same form of representation (e.g., Noel & Seron, 1993; Dehaene & Akhavein, 1995), the target, which was an external Arabic numeral, and the standard, which was an internal number word, must be converted into a common representation. This common representation could be a phonological representation of number words or, as proposed by Dehaene, Dupoux, and Mehler (1990) and Hinrichs, Yurko, and Hu (1981), a format-independent line-like analog representation. The result was a holistic comparison.

In our present study, both the target and the standard were already in the same form of representation: they were both external Arabic numerals. They need not be converted into another form of representation before they could be compared. The dimensional structures of Arabic numerals could directly activate a set of perceptual and cognitive processes. The power dimension could activate perceptual processes such as perceptual identification of the positions of individual digits and perceptual search of specific digits, and the base dimension could activate cognitive processes such as memorial retrieval of the values of individual digits and mental comparison of individual digits. These perceptual and cognitive processes interacted with each other and worked together to execute the comparison task. Both the perceptual and the cognitive processes were the basic processes for the comparison. Neither was a peripheral process for the other. The result was a parallel comparison, caused by the interaction of the perceptual and cognitive processes in the form of a Stroop-like effect.

Theories of Number Representations

Our present study shows that under different representations the same two-digit number comparison task generated different results. This is an example of the representational effect. In general, different numerals (e.g., Arabic, Roman, Greek, Chinese, etc.), though they all represent the same abstract structure (numbers), can often have different representational efficiencies and cause dramatically different behaviors, as can be seen from the difficulty difference between 73×27 (Arabic numerals) and $LXXIII \times XXVII$ (Roman numerals). One of the central issues in numerical cognition is whether the internal representations of different types of numerals have separate, format-specific representations or a single, format-independent representation.

Format-Independent Representations

As discussed at the beginning of this article, the distance effect for one-digit Arabic numerals also exists for English written number words, Japanese Kanji and kana numerals, line lengths, and dot patterns. This suggests that different numerals (at least one-digit numerals) may have a format independent, line-like analog internal representation. The format-independent view was explicitly formulated by McCloskey and his colleagues (McCloskey, 1992; McCloskey, Caramazza, & Basili, 1985; McCloskey, Sokol, & Goodman, 1986). Based on the studies of the dissociation of different numerical processing stages in dyscalculia, they proposed a model that assumes a single abstract representation for all types of numerals and number words. This single abstract representation links three functionally independent modules for number comprehension, calculation, and production, which have no direct links with each other. The comprehension and production modules carry out transformations (encoding and decoding) between external representations and the single abstract internal representation. The calculation module, which contains cognitive processes and knowledge of arithmetic, only operates upon the abstract representation. According to this model, the

representational effect only occurs at the stages of comprehension and production, which have different processes for different external number representations. For example, the representational effect caused by different types of number words (preferred and non-preferred languages) in bilinguals (March & Maki, 1976; McClain & Huang, 1982) can be explained in terms of the different comprehension and production processes for different types of number words.

Format-Specific Representations

According to the format-specific view, the representational effect in numerical tasks is caused by not just the different comprehension and production processes but also the different internal representations that are specific to different external representations. For example, the studies by Gonzalez & Kolers (1982; 1987) suggest that Arabic and Roman numerals were processed differently. They argue that during numerical processing people do not transform different external number representations into a common abstract internal representation. Rather, they operate upon different internal representations that reflect the physical characteristics of different external representations. Deloche and Seron (1987) also suggested a transcoding mechanism that does not require a central abstract representation. In this view, one type of number representations (e.g., Arabic numerals) can be directly transcribed into another type of number representations (e.g., English number words) without the mediation of any central representations. As a representative for the format-specific view, the encoding-complex model developed by Campbell and Clark (1988; Clark & Campbell, 1991) denies any central abstract representations. In this model, different forms of external number representations activate different forms of internal number representations, which are functionally integrated in an encoding complex. Within the complex, one form of internal representations can potentially activate other forms of internal representations,

and any form of internal representations can be involved in the comprehension, calculation, and production of numbers.

Distributed Representations

Both the format-independent and the format-specific views are only concerned with internal representations of numbers: how people perform numerical tasks in their heads, how numbers and arithmetic facts are represented in memory, and what mental processes and procedures are involved in the comprehension, calculation, and production of numbers. In addition, both views are developed mainly from the studies of simple numerical tasks that can be performed entirely in internal representations, such as single or double-digit number comparison and simple arithmetic. Internal number representations are certainly important. However, the theories and models developed for them can not account for numerical tasks that involve external representations. As an example of numerical tasks that involve external representations but are possible to perform without them, the line-like analog representation assumed for two-digit Arabic numerals can only account for the holistic comparison when the comparison is performed on internal representations but cannot account for the parallel comparison when the comparison is performed on external representations. As an example of numerical tasks that are impossible to perform without external representations, consider the task of multiplying 735 by 278. Without special training and expertise, nobody can perform this task by encoding the two three-digit numerals in internal representations and carrying out the calculation entirely in internal representations.

Therefore, to account for the full spectrum of numerical tasks, including those involving only internal representations and those involving both internal and external representations, we should consider number representations in general as distributed representations that have internal and external representations as two indispensable components. We should also consider numerical tasks as distributed cognitive tasks that

require the interactive processing of information distributed across internal and external representations (see Zhang & Norman, 1994, 1995). There are not yet enough empirical data for us to propose a detailed process model with explicitly specified processing mechanisms for external representation based numerical tasks. Nevertheless, we can specify a set of properties of such tasks that should be considered by any potential process models. First, external representation based numerical tasks involve interactive, integrative, and dynamic processing of information perceived from external representations and that retrieved from internal representations. Second, the processes in such tasks are activated and determined by representations, not vice versa. Thus, even if different types of numerals (Arabic, Roman, etc.) might have a single internal representation, they have different distributed representations because their external representations are different. Therefore, it is different distributed representations that are responsible for the representational effect in numerical tasks. For number comparison, this means that there is no single comparison process for all types of number representations. Fourth, external representations need not be re-represented as internal representations in order to be involved in numerical tasks: they can directly activate perceptual processes and directly provide perceptual information that, in conjunction with the memorial information and cognitive processes provided by internal representations, determine the behavior in numerical tasks. Perceptual processes are basic processes of numerical tasks, just like cognitive processes. They directly operate upon external representations to participate in numerical tasks.

Conclusion

The present study explored the effect of external representations on number comparison tasks. A few trivial modifications on the previously reported two-digit number comparison task produced completely different results. The difference in results between previous and current studies is a reflection of representational forms: the

comparison was based on internal representations for previous studies but on external representations for our present study. This effect on number comparison caused by external representations supports the framework of distributed number presentations. In complex numerical tasks that involve external representations, number representations should be considered as distributed representations and the behavior in these tasks should be considered as the interactive processing of internal and external information through the interplay of perceptual and cognitive processes. We suggest that any general theory of number representations and any process models of numerical tasks should consider external representations as an essential component.

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second International Symposium on Information Theory* (pp. 267). Budapest: Akademiai Kiado.
- Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute*, 50, 277-290.
- Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.
- Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, 103, 1131-1136.
- Campbell, J. I. D., & Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General*, 117 (2), 204-214.
- Clark, J. M., & Campbell, J. I. D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition*, 17(2), 204-239.
- Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics*, 45 (6), 557-566.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1-42.
- Dehaene, S. (1993). *Numerical cognition*. Cambridge, MA: Blackwell Publishers.
- Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (2), 314-326.
- Dehaene, S., Dupoux, E. & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 626-641.

- Deloche, G., & Seron, X. (1987). Numerical transcoding: A general production Model. In G. Deloche & X. Seron (Ed.), *Mathematical disabilities: A cognitive neuropsychological perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 442-453.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43-74.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, Md., Lawrence Erlbaum Associates.
- Gonzalez, E. G., & Kolers, P. A. (1982). Mental manipulation of arithmetic symbols. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8* (4), 308-319.
- Gonzalez, E. G., & Kolers, P. A. (1987). Notational constraints on mental operations. In G. Deloche & X. Seron (Ed.), *Mathematical disabilities: A cognitive neuropsychological perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hinrichs, J. V., Berie, J. L., & Mosell, M. K. (1982). Place information in multidigit number comparison. *Memory and Cognition*, *10*, 487-495.
- Hinrichs, J. V., Yurko, D. S. & Hu, J. M. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 890-901.
- Marsh, L. G., & Maki, R. H. (1976). Efficiency of arithmetic operations in bilinguals as a function of language. *Memory & Cognition*, *4* (4), 459-464.
- McClain, L., & Huang, J. Y. S. (1982). Speed of simple arithmetic in bilinguals. *Memory & Cognition*, *10* (6), 591-596.
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107-157.

- McCloskey, M., Caramazza, A., & Basili, A. G. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*, 171-196.
- McCloskey, M., Sokol, S. M., & Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General, 115*, 307-330.
- Moyer, R. S., & Dumais, S. T. (1978). Mental comparison. In G. Bower (Ed.), *The Psychology of learning and motivation*. New York: Academic Press.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519-1520.
- Myung, I. J., & Pitt, M. A. (in press). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 00*, 000-000.
- Noel, M. P., & Seron, X. (1993). Arabic number reading deficit: A single case study. *Cognitive Neuropsychology, 10*, 317-339.
- Pollock, S. E., & Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 32-45.
- Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83*, 274-278.
- Takahashi, A., & Green, D. (1983). Numerical judgments with kanji and kana. *Neuropsychologia, 21*, 259-263.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.
- Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.
- Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition, 57*, 271-295.

Zhang, J. (in press). The nature of external representations in problem solving.
Cognitive Science, 00, 000-000.

AUTHOR NOTES

This research was in part supported by Grants N00014-95-1-0241 and N00014-96-1-0472 from the Office of Naval Research, Cognitive and Neural Sciences Technology Division, and by a Seed Grant from the Office of Research at The Ohio State University. We would like to thank Gwen Hall for his assistance in the experiments. Correspondence and request for reprints should be sent to Jiajie Zhang, Department of Psychology, The Ohio State University, 1827 Neil Avenue, Columbus, Ohio 43210, USA. Email: zhang.52@osu.edu.

FOOTNOTES

¹ These two models of Macintosh computers could only provide a 16 ms resolution. The 1 ms resolution was achieved by using the millisecond timer software developed by Dan Costin.

² $AIC_i = -2\ln(ML_i) + 2n_i$, where ML_i is the maximum likelihood for Model i and n_i is the number of free parameters in the model. The criterion prescribes that the model that minimizes the AIC should be chosen. AIC can be rewritten as a function of SSE (sum squared error): $AIC_i = N \ln(SSE_i) + 2n_i + (-N \ln N + \frac{1}{N} + N \ln 2p)$, where SSE_i is the Sum Squared Error for Model i , n_i is the number of free parameters in the model, and N is the number of observations (Myung, personal communication). The latter equation was used in the present study.

Table 1: AIC values of Model Fitting (R^2 in parentheses)

		Standard 65		Standard 55	
Models		Smaller than	Larger than	Smaller than	Larger than
		Standard	Standard	Standard	Standard
Experiment 1	Sequential	342 (0.49)	280 (0.88)	401 (0.67)	361 (0.90)
	Parallel	311 (0.81)	291 (0.84)	368 (0.85)	356 (0.92)
	Holistic	322 (0.70)	309 (0.70)	375 (0.81)	358 (0.91)
Experiment 2	Sequential	338 (0.63)	304 (0.66)	394 (0.72)	383 (0.61)
	Parallel	317 (0.81)	306 (0.66)	370 (0.84)	376 (0.66)
	Holistic	332 (0.68)	325 (0.34)	372 (0.83)	377 (0.65)
Experiment 3	Sequential	287 (0.73)	263 (0.84)	356 (0.80)	376 (0.67)
	Parallel	284 (0.76)	255 (0.88)	347 (0.85)	369 (0.74)
	Holistic	283 (0.74)	280 (0.72)	358 (0.78)	367 (0.72)

FIGURE CAPTIONS

Figure 1. Three models of two-digit number comparison. Each graph represents the reaction times of comparing target numerals 11-54 and 56-99 with the standard 55. See Equations 1, 2, and 3 and the text for detailed explanations.

Figure 2. Reaction times for targets compared with 65 in Experiment 1.

Figure 3. Reaction times for targets compared with 55 in Experiment 1.

Figure 4. Unit effect for standards 65 and 55 in Experiment 1. The RT for each target in a decade was subtracted by the mean RT of the corresponding decade, then averaged across decades 30s, 40s, and 50s and across 70s, 80s, and 90s for standard 65 and averaged across 10s, 20s, 30s, and 40s and across 60s, 70s, 80s, and 90s for standard 55.

Figure 5. Reaction times for targets compared with 65 in Experiment 2.

Figure 6. Reaction times for targets compared with 55 in Experiment 2.

Figure 7. Unit effect for standards 65 and 55 in Experiment 2. The RT for each target in a decade was subtracted by the mean RT of the corresponding decade, then averaged across decades 30s, 40s, and 50s and across 70s, 80s, and 90s for standard 65 and averaged across 10s, 20s, 30s, and 40s and across 60s, 70s, 80s, and 90s for standard 55.

Figure 8. Reaction times for targets compared with 65 in Experiment 3.

Figure 9. Reaction times for targets compared with 55 in Experiment 3.

Figure 10. Unit effect for standards 65 and 55 in Experiment 3. The RT for each target in a decade was subtracted by the mean RT of the corresponding decade, then averaged across decades 30s, 40s, and 50s and across 70s, 80s, and 90s for standard 65 and averaged across 10s, 20s, 30s, and 40s and across 60s, 70s, 80s, and 90s for standard 55.

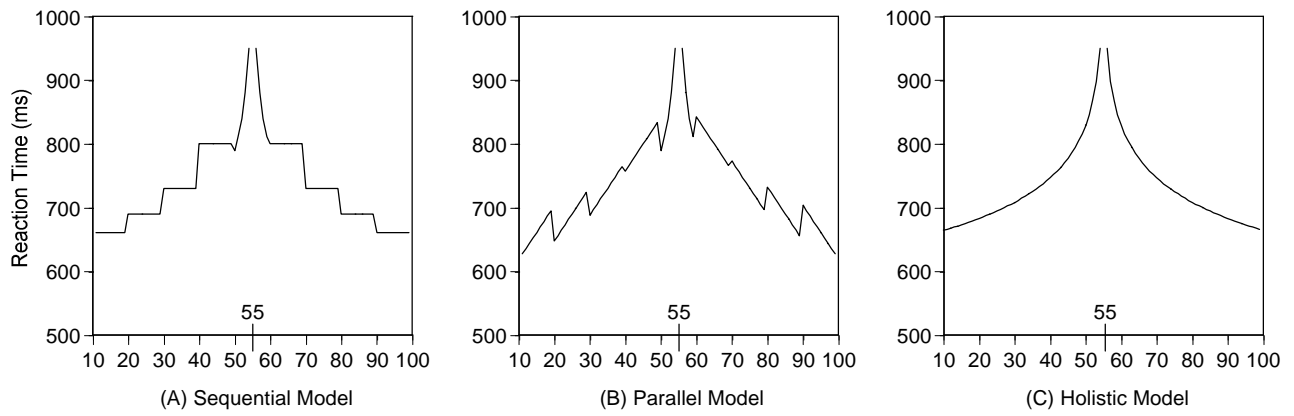


Figure 1

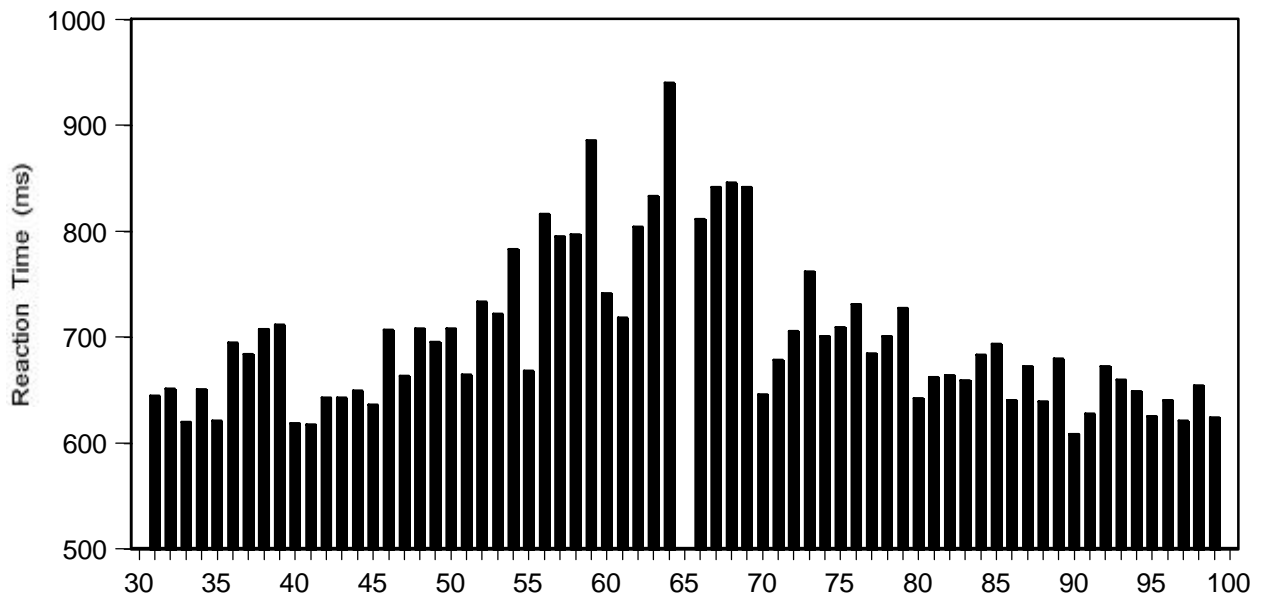


Figure 2

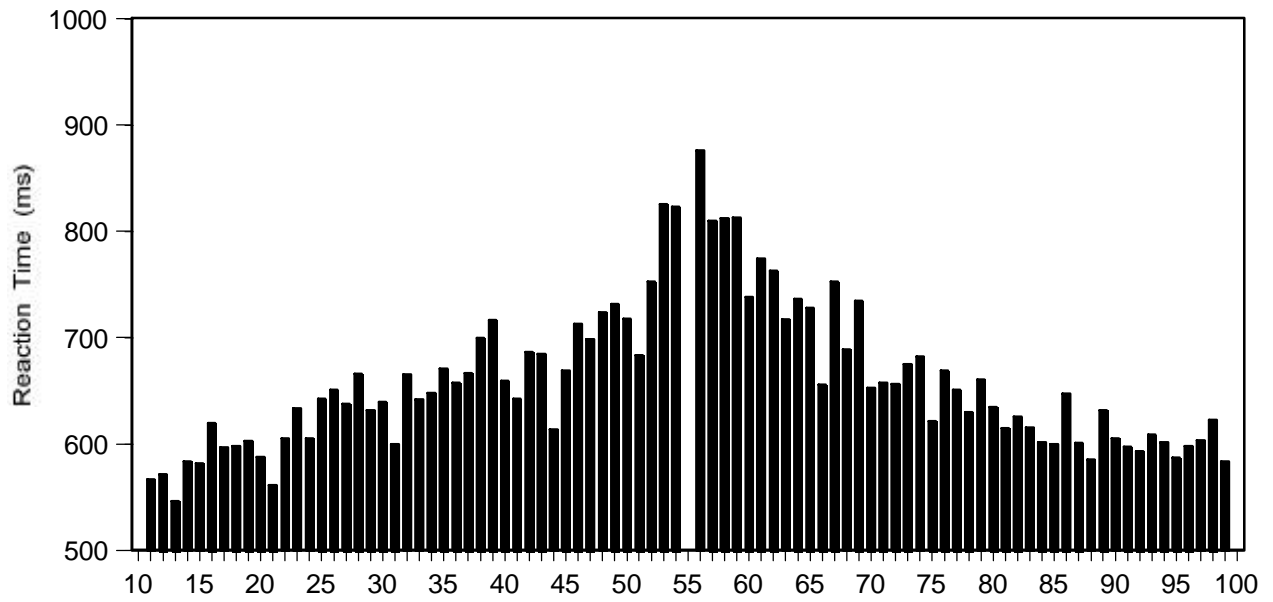
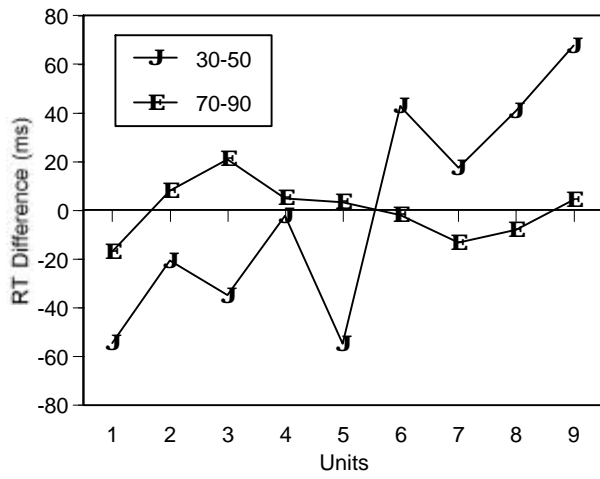
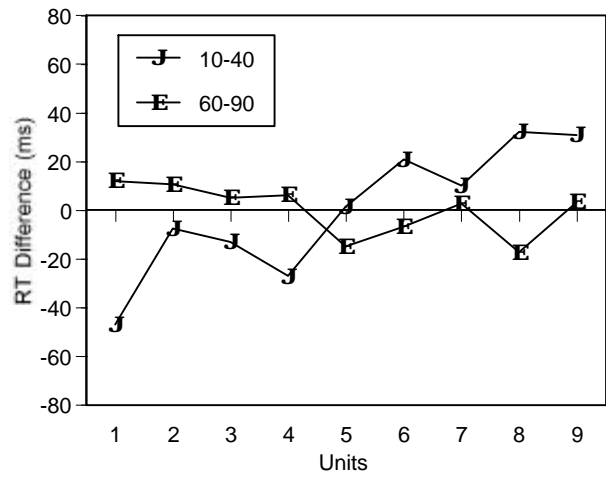


Figure 3



(A) Standard 65



(B) Standard 55

Figure 4

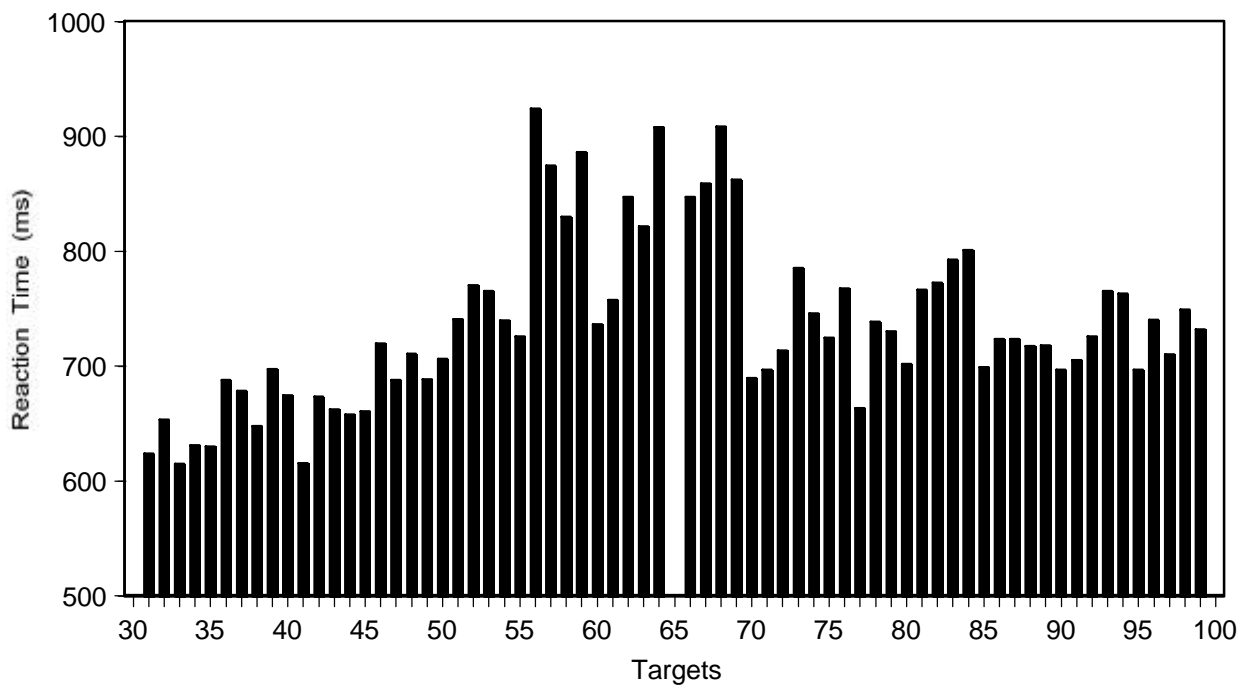


Figure 5

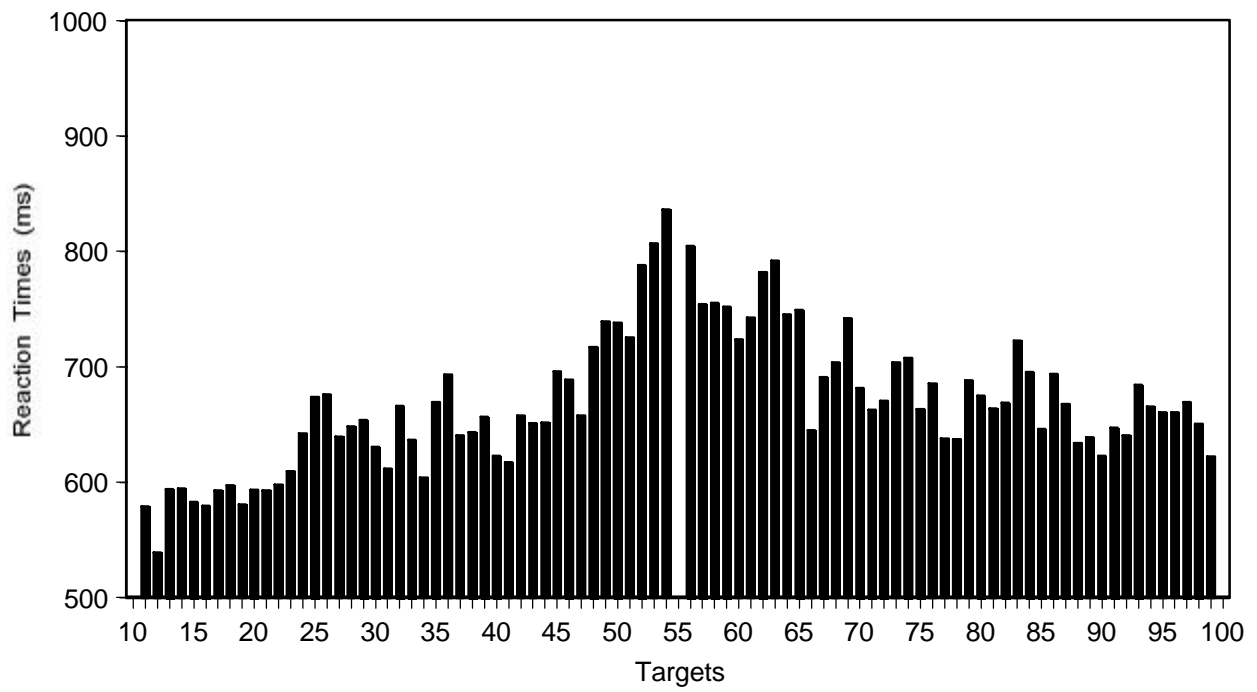
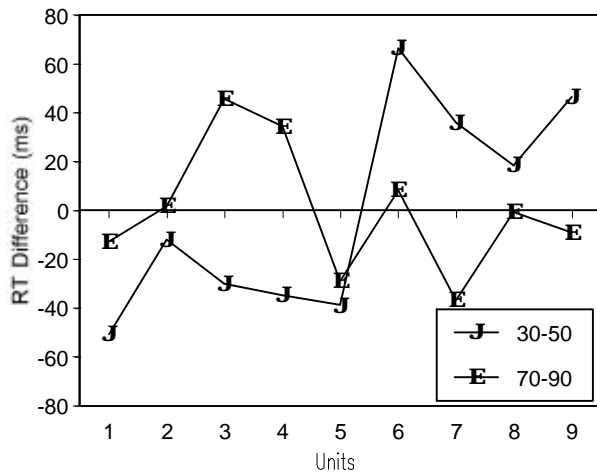
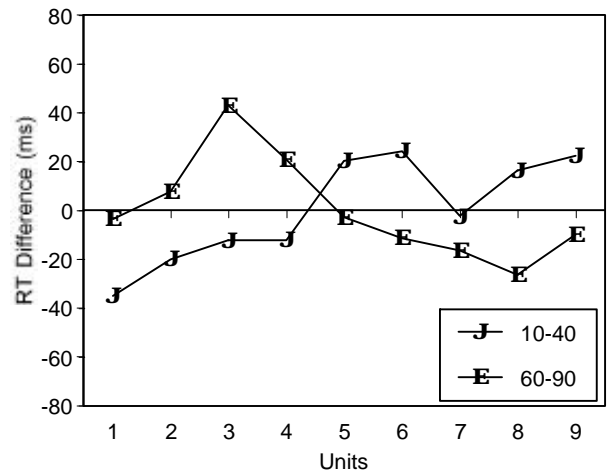


Figure 6



(A) Standard 65



(B) Standard 55

Figure 7

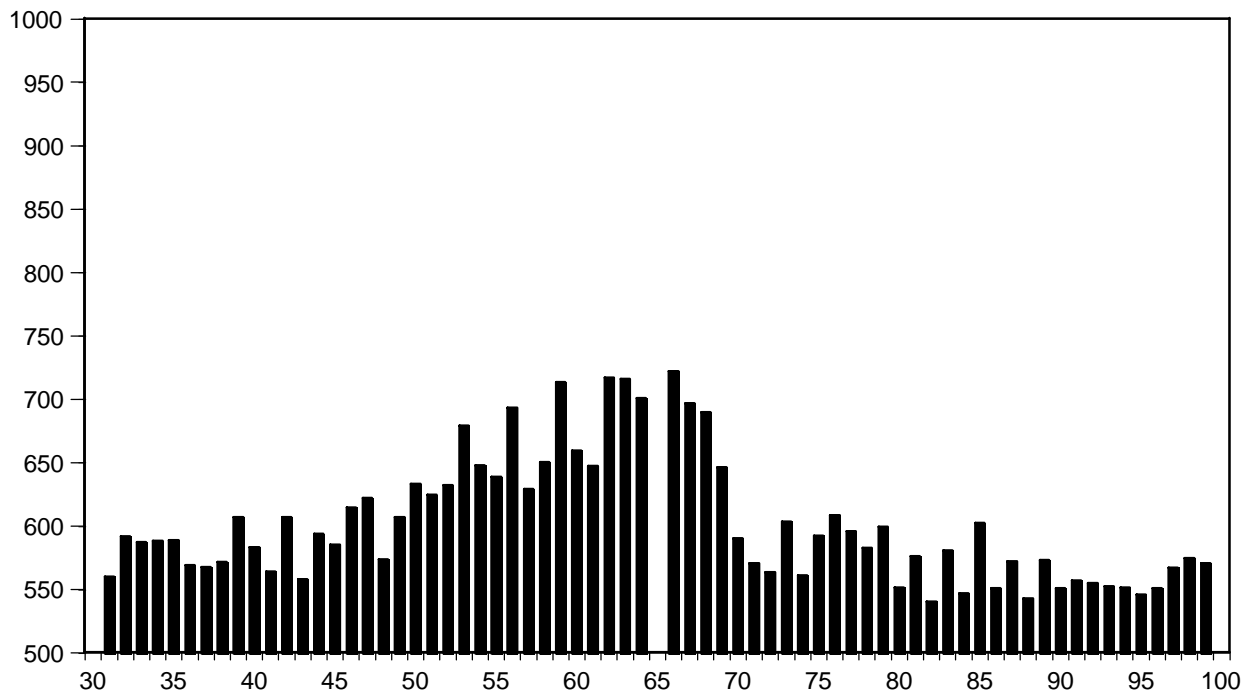


Figure 8

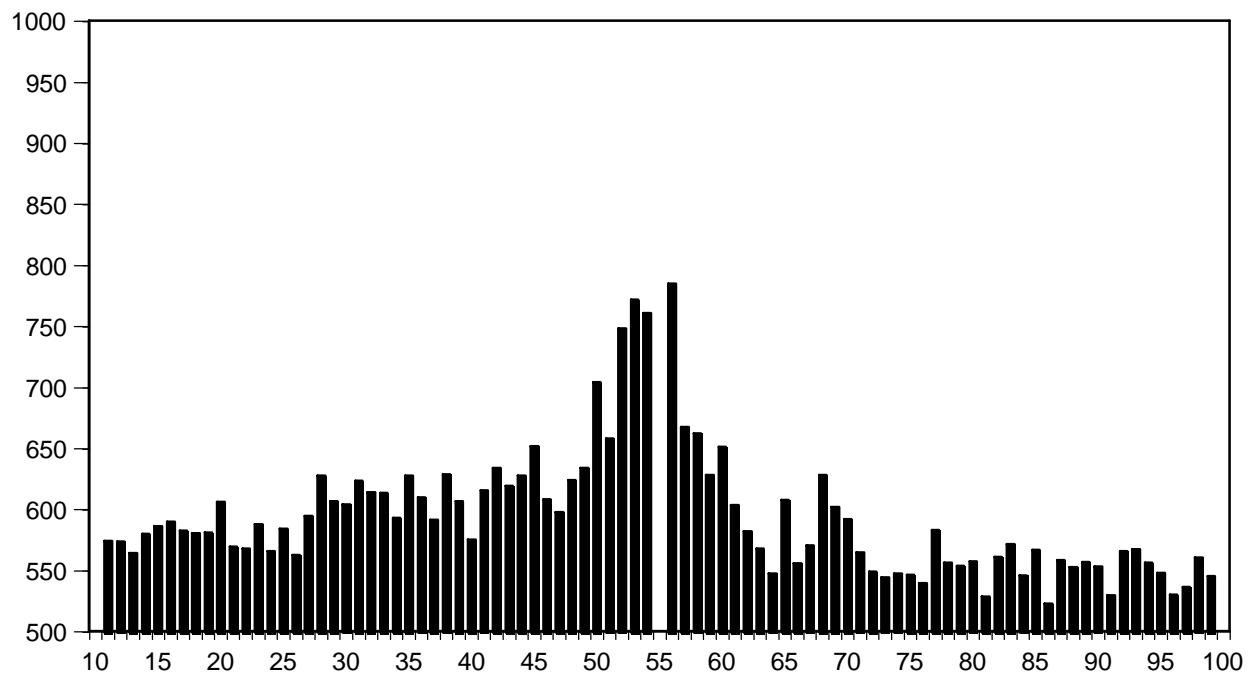
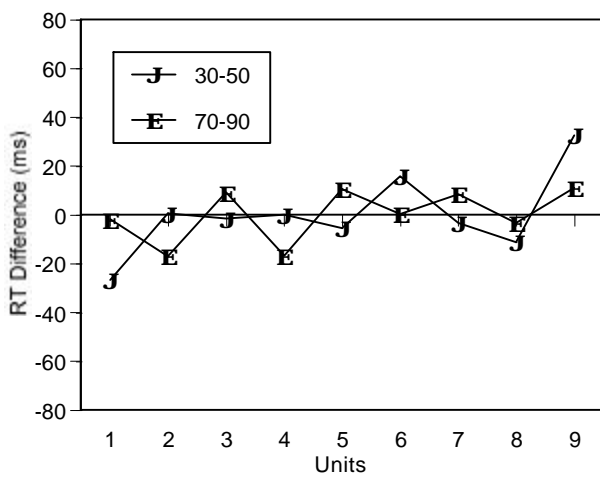
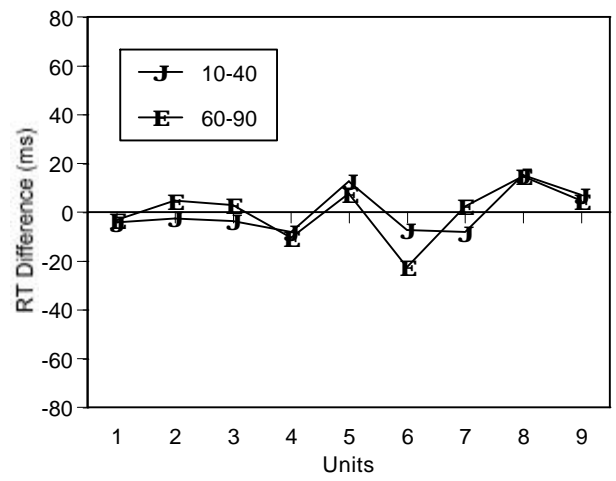


Figure 9



(A) Standard 65



(B) Standard 55

Figure 10